Retention Stage: Customer Lifetime Value (CLTV) Prediction Model

Customer Lifetime Value (CLTV) prediction is a critical aspect of customer retention strategy. It helps in identifying the long-term value of a customer to the bank, allowing for strategic decisions on where to invest marketing and retention efforts.

Customer Lifetime Value (CLTV) Prediction Model

Objective:

- Predict the lifetime value of customers to prioritize high-value customers for retention and targeted marketing campaigns.
- Optimize resource allocation by focusing on customers with the highest potential value.

Methodology:

1. **Data Preparation:**
    - Collect and preprocess customer data, including transactional, behavioral, and demographic information.
    - Historical transaction data and customer interaction logs are crucial.
2. **Attribute Selection:**
    - **Raw Attributes:**
        - Customer ID
        - Age
        - Income
        - Occupation
        - Credit score (CIBIL score)
        - Product holdings (e.g., savings account, loan, credit card)
        - Transaction history (frequency, recency, monetary value)

    - Customer interactions (e.g., website visits, customer service interactions)
    - Complaint logs and resolution times
    - Tenure with the bank
  - **Derived Attributes:**
    - Average purchase value: Average value of transactions.
    - Purchase frequency: Number of purchases in a given period.
    - Recency: Time since the last purchase.
    - Engagement score: Measure of overall engagement with the bank's services.
    - Customer satisfaction score: Derived from surveys or feedback forms.
    - Churn probability: Probability of the customer leaving the bank.
  - **Target Variable:**
    - CLTV: Predicted monetary value of a customer over their entire relationship with the bank.
3. **Model Selection:**
  - **Regression Algorithms:**
    - Linear Regression
    - Random Forest Regressor
    - Gradient Boosting Regressor
    - XGBoost
    - Neural Networks
  - **Evaluation Metrics:**
    - Mean Absolute Error (MAE)
    - Mean Squared Error (MSE)
    - Root Mean Squared Error (RMSE)
    - R-squared (R^2)

Implementation Steps

1. **Data Collection and Preparation:**

```
import pandas as pd
from sklearn.preprocessing import StandardScaler, OneHot
Encoder
from sklearn.model_selection import train_test_split

# Sample customer data
data = {
    'customer_id': [1, 2, 3, 4, 5],
    'age': [25, 45, 35, 50, 23],
    'income': [50000, 120000, 75000, 100000, 45000],
    'cibil_score': [700, 800, 750, 780, 690],
    'product_holdings': ['savings,loan', 'savings,credit
card', 'savings,loan', 'savings', 'loan'],
    'transaction_frequency': [15, 50, 25, 30, 10],
    'average_transaction_value': [2000, 5000, 3000, 400
0, 1500],
    'engagement_score': [3, 5, 4, 4, 2],
    'tenure': [2, 10, 5, 8, 1],
    'cltv': [20000, 120000, 75000, 100000, 45000]
}
df = pd.DataFrame(data)

# Encode categorical variables
encoder = OneHotEncoder()
encoded_products = encoder.fit_transform(df[['product_ho
ldings']]).toarray()
df = df.join(pd.DataFrame(encoded_products, columns=enco
der.get_feature_names_out(['product_holdings']))).drop
('product_holdings', axis=1)

# Standardize numerical features
scaler = StandardScaler()
numerical_features = ['age', 'income', 'cibil_score', 't
ransaction_frequency', 'average_transaction_value', 'eng
agement_score', 'tenure']
```

```
df[numerical_features] = scaler.fit_transform(df[numeric
al_features])

# Split data into train and test sets
X = df.drop(['customer_id', 'cltv'], axis=1)
y = df['cltv']
X_train, X_test, y_train, y_test = train_test_split(X,
y, test_size=0.2, random_state=42)
```

2. **Model Training and Evaluation:**

   Gradient Boosting Regressor:

```
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.metrics import mean_absolute_error, mean_sq
uared_error, r2_score
import numpy as np

# Train Gradient Boosting model
model = GradientBoostingRegressor(n_estimators=100, rand
om_state=42)
model.fit(X_train, y_train)

# Predict on test data
y_pred = model.predict(X_test)

# Evaluate model performance
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
r2 = r2_score(y_test, y_pred)

print(f'Mean Absolute Error: {mae}')
print(f'Mean Squared Error: {mse}')
print(f'Root Mean Squared Error: {rmse}')
print(f'R-squared: {r2}')
```

Variables for CLTV Prediction

1. **Customer Attributes:**
    - **Customer ID:** Unique identifier for each customer.
    - **Age:** Different age groups may have varying lifetime values.
    - **Income:** Financial stability can impact spending and saving behaviors.
    - **Occupation:** Reflects lifestyle and financial needs.
    - **Credit Score:** Higher scores may correlate with higher lifetime values.
    - **Product Holdings:** Number and type of products owned by the customer.
2. **Transactional and Behavioral Data:**
    - **Transaction Frequency:** Frequency of transactions can indicate engagement.
    - **Average Transaction Value:** Monetary value of transactions.
    - **Engagement Score:** Overall interaction and activity with the bank.
    - **Tenure:** Duration of the customer's relationship with the bank.
    - **Complaint Logs:** History of customer complaints and resolution times.
3. **Derived Attributes:**
    - **Average Purchase Value:** Average value of transactions.
    - **Purchase Frequency:** Number of purchases in a given period.
    - **Recency:** Time since the last purchase.
    - **Engagement Score:** Comprehensive measure of customer activity.
    - **Customer Satisfaction Score:** Derived from feedback or survey responses.
    - **Churn Probability:** Probability of the customer leaving the bank.
4. **Target Variable:**
    - **CLTV:** Predicted monetary value of a customer over their entire relationship with the bank.

Conclusion

By implementing a CLTV prediction model, the bank can effectively identify high-value customers and focus its retention and marketing efforts on them.

This strategic approach ensures that resources are allocated efficiently, maximizing long-term profitability and enhancing customer satisfaction. Using advanced regression techniques and a comprehensive set of attributes, the bank can accurately predict customer lifetime value and make informed business decisions.